33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Learning Event-Related Potentials (ERPs) from Multichannel EEG Recordings: A Spatio-Temporal Modeling Framework with a Fast Estimation Algorithm

Wei Wu[1], *Student Member, IEEE*, Shangkai Gao[1], *Fellow, IEEE*

Abstract— **Extracting event-related potentials (ERPs) from multichannel EEG recordings remains a challenge due to the poor signal-to-noise ratio (SNR). This paper presents a multivariate statistical model of ERPs by exploiting the existing knowledge about their spatio-temporal properties. In particular, a computationally efficient algorithm is derived for fast model estimation. The algorithm, termed SIM, can be intuitively interpreted as maximizing the signal-to-noise ratio in the source space. Using both simulated and real EEG data, we show that the algorithm achieves excellent estimation performance and substantially outperforms a state-of-the-arts algorithm in classification accuracies in a P300 target detection task. The results demonstrate that the proposed modeling framework offers a powerful tool for exploring the spatio-temporal patterns of ERPs as well as learning spatial filters for decoding brain states.**

I. INTRODUCTION

Event-related potentials (ERPs) are brain responses both time-locked and phase-locked to external stimuli [1]. They can serve as "windows" on the brain and mind for studying cognitive functions. However, since the ERP (signal) is typically embedded in strong stimulus-unrelated spontaneous EEG activities (noise), enhancing the signal-to-noise ratio (SNR) entails averaging EEG data over a large number of trials. Thanks to the advances in the high-density EEG recording technique in the past decades, efforts have been made to enable single-trial analysis of the ERP by exploiting the spatial information of multichannel EEG signals [2]. The spatial information is typically extracted in the form of spatial filters at the training stage based on multiple-trial EEG.

A variety of techniques have been proposed for optimizing spatial filters for ERPs, with varying degrees of success. These encompass principal component analysis (PCA) [4], independent component analysis (ICA) [5], and sparse component analysis (SCA) [6]. A drawback with these techniques is that they are not designed specifically for extracting ERPs, hence achieving only suboptimal performance. As a supervised variant of ICA, a *regularized* second-order blind identification (SOBI) algorithm was proposed in [7] by biasing the extraction focus of the SOBI algorithm towards the subspace of the phase-locked components. It was shown that substantial performance gain in the SNR was attained over the unregularized SOBI algorithm. Nonetheless, the algorithm needs to be run multiple times to determine the optimal degree of regularization, resulting in high computational cost. Moreover, the regularization scheme is based largely on heuristics, for which there is no guarantee that the resulting ERP sources attain the highest SNRs. Recently, another spatio-temporal filtering method was proposed in [8] as a noise canceller in the spatial domain. However, to constrain the extracted ERP sources the method requires predefined templates of the ERP waveforms, which can only be obtained in an ad-hoc manner in many situations and the choice often seems arbitrary.

In pursuing "optimal" spatial filters for retrieving ERPs, ideally one would like them to maximize the ERP power while being maximally orthogonal to spontaneous activities. Besides, from a practical viewpoint, to be applicable in real-time decoding of brain states the learning procedure should possess fast computational speed to adapt to brain dynamics. Inspired by these considerations, in this paper we contribute a spatio-temporal modeling framework for learning ERPs from multichannel EEG recordings. A fast algorithm, which specifically maximizes the SNR, is derived for efficient spatial filter design. The efficacy of the algorithm is demonstrated via the analysis of both simulated and real EEG data.

II. A SPATIO-TEMPORAL MODEL OF ERPS

A. Motivation

Our proposed spatio-temporal model of ERPs is motivated by two observations:

- According to previous neurophysiological studies [1], the ERP wave is *approximately* identical under repetitive stimuli.
- The spontaneous EEG activities can be approximately modeled by Gaussian distributions.

The first observation derives from the phase locking property of ERPs. Although in practice there may well be inter-trial variability in ERP amplitude and latency, the variations are typically small compared with spontaneous activities. In this case, the observation can still lead to a *useful* model [2]. The second observation can be understood by referring to the fact that the Gaussianity assumption is suited to modeling mildly amplitude-modulated oscillatory activities as their kurtoses are close to zero, a hallmark of Gaussian random variables [3].

[1]Department of Biomedical Engineering, Tsinghua University, Beijing 100084, China. Email: weiwu@neurostat.mit.edu. This work was supported by National Natural Science Foundation of China under Grant 90820304, and National Basic Research Program (973) of China (No. 2011CB933204).

B. Model Formulation

Let c, m, n, k ($c \in [1, C], m \in [1, M], n \in [1, N], k \in [1, K]$) denote the indices for channels, sources, sampled time points, and trials, respectively. Within a specific trial k, our proposed overdetermined (i.e., $M \le C$) spatio-temporal model of ERPs is posed as follows:

$$\mathbf{x}_{kn} = \mathbf{A}\mathbf{z}_n + \mathbf{e}_{kn}, \quad n = 1, 2, \cdots, N \tag{1}$$

where $\mathbf{x}_{kn} \in \mathbb{R}^C$ denotes the multichannel EEG signals; $\mathbf{z}_n \in \mathbb{R}^M$ denotes the ERP source signals; $\mathbf{e}_{kn} \in \mathbb{R}^C$ denotes the noise term consisting of spontaneous EEG activities as well as any other activities that are uncorrelated with ERP. $\mathbf{A} \in \mathbb{R}^{C\times M}$ is the mixing matrix relating in a linear fashion (up to the additive noise) the set of M ERP sources to the set of C-channel EEG signals on the scalp. $\mathbf{e}_{kn}$ is modeled by a zero-mean multivariate Gaussian distribution with the covariance matrix being $\mathbf{\Psi}$:

$$p(\mathbf{e}_{kn}) = \mathcal{N}(\mathbf{e}_{kn}|\mathbf{0}, \mathbf{\Psi})$$

How are the aforementioned two observations manifested in model (1)? The first observation is enforced by the assumption that the time course of each ERP source remains identical across trials. Note that $\mathbf{z}_n$ ($n = 1, 2, \cdots, N$) are treated as *deterministic parameters* in the model; no specific distributional assumption is imposed on them. Besides, the separate ERP sources allow for the modeling of ERP subcomponents that may reflect functionally distinct cognitive processes. The second observation is taken into account by the Gaussian distribution assumed for $\mathbf{e}_{kn}$. Note that in this case the covariance matrix $\mathbf{\Psi}$ is unlikely to be diagonal.

C. Estimation Algorithm

The log-likelihood function for model (1) is

$$L = -\frac{N}{2}\sum_{k=1}^{K}\Big[N\ln(2\pi) + \ln|\mathbf{\Psi}| + \tag{2}$$

$$\frac{1}{N}\mathrm{Tr}\Big\{\mathbf{\Psi}^{-1}(\mathbf{X}_k - \mathbf{A}\mathbf{Z})(\mathbf{X}_k - \mathbf{A}\mathbf{Z})^T\Big\}\Big] \tag{3}$$

where $\mathbf{X}_k = [\mathbf{x}_{k1}, \mathbf{x}_{k2}, \cdots, \mathbf{x}_{kN}]$, $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_N]$, $\mathrm{Tr}\{\cdot\}$ denotes the trace of a matrix.

The goal is to obtain the maximum likelihood estimates (MLEs) of the parameters $\{\mathbf{A}, \mathbf{Z}, \mathbf{\Psi}\}$ from the training data. Suppose $\mathbf{\Psi}$ is known, we have the following theorem[1]:

Theorem 1: The MLEs of $\mathbf{A}$ and $\mathbf{Z}$ in model (1), denoted by $\mathbf{A}_{\mathrm{ML}}$ and $\mathbf{Z}_{\mathrm{ML}}$ respectively, are solutions to the following optimization problem:

$$\min_{\mathbf{A},\mathbf{Z}} \|\mathbf{\Psi}^{-1/2}\bar{\mathbf{X}} - \mathbf{\Psi}^{-1/2}\mathbf{A}\mathbf{Z}\|_F \tag{4}$$

where $\|\mathbf{B}\|_F$ denotes the Frobenius norm of matrix $\mathbf{B}$, $\bar{\mathbf{X}} = \frac{1}{K}\sum_{k=1}^{K}\mathbf{X}_k$ is the EEG data matrix averaged across trials, $\mathbf{\Psi}^{-1/2}$ denotes the inverse of the principal square root of $\mathbf{\Psi}$.

Since $\mathbf{\Psi}^{-1/2}$ whitens the noise term $\mathbf{e}_{kn}$ in model (1), Problem (4) is a low-rank matrix approximation problem in the whitened space. As such, for notational convenience henceforth $\mathbf{\Psi}^{-1/2}\mathbf{A}$ and $\mathbf{\Psi}^{-1/2}\bar{\mathbf{X}}$ are denoted by $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{X}}$, respectively. It can be shown that $\tilde{\mathbf{A}}_{\mathrm{ML}}\mathbf{Z}_{\mathrm{ML}}$ is given by the leading M factors of the SVD of $\tilde{\mathbf{X}}$. However, to further identify $\tilde{\mathbf{A}}_{\mathrm{ML}}$ and $\mathbf{Z}_{\mathrm{ML}}$ we notice that there is the issue of multiplication indeterminancy, i.e., $\tilde{\mathbf{A}}_{\mathrm{ML}}$ can be right-multiplied by any invertible matrix $\mathbf{P}$ as long as $\mathbf{Z}_{\mathrm{ML}}$ is left-multiplied by $\mathbf{P}^{-1}$.

To ensure uniqueness, analogous to the basic idea of PCA we assume that $\tilde{\mathbf{A}}_{\mathrm{ML}}$ is an orthogonal matrix and seek $\tilde{\mathbf{A}}_{\mathrm{ML}}$ such that the resulting $\mathbf{Z}_{\mathrm{ML}}$ yields ERP sources with the maximum power in the source space[2]:

$$\max_{\tilde{\mathbf{A}}^+} \mathrm{Tr}\Big\{\tilde{\mathbf{A}}^+\mathbf{R}_s[\tilde{\mathbf{A}}^+]^T\Big\} \quad s.t. \;\; \tilde{\mathbf{A}}^+[\tilde{\mathbf{A}}^+]^T = \mathbf{I} \tag{5}$$

where $\mathbf{R}_s = \frac{1}{N}\tilde{\mathbf{X}}\tilde{\mathbf{X}}^T$, $\tilde{\mathbf{A}}^+$ denotes the Penrose-Moore pseudo-inverse of $\tilde{\mathbf{A}}$.

From a pattern recognition's viewpoint, maximizing the power of the ERP sources can desirably increase the sensitivity of the ERP detection. Moreover, since the noise term has been normalized, the objective function in Problem (5) can also be viewed as the SNR of the ERPs in the source space. Thus the foregoing estimation procedure can be interpreted as an SNR maximizer, with $\tilde{\mathbf{A}}^+_{\mathrm{ML}}$ acting as the *spatial filters*.

Problem (5) is known to be solved by the eigenvalue decomposition of $\mathbf{R}_s$, for which many scientific computing softwares offer standard routines for fast computation (e.g., eig.m in MATLAB). Once $\mathbf{A}_{\mathrm{ML}} = \mathbf{\Psi}^{1/2}\big[\tilde{\mathbf{A}}^+_{\mathrm{ML}}\big]^+$ and $\mathbf{Z}_{\mathrm{ML}} = \tilde{\mathbf{A}}^+_{\mathrm{ML}}\tilde{\mathbf{X}}$ are obtained, the MLE of $\mathbf{\Psi}$ i.e., $\mathbf{\Psi}_{\mathrm{ML}}$, in model (1) can be updated as follows:

$$\mathbf{\Psi}_{\mathrm{ML}} = \frac{1}{KN}\sum_{k=1}^{K}(\mathbf{X}_k - \mathbf{A}_{\mathrm{ML}}\mathbf{Z}_{\mathrm{ML}})(\mathbf{X}_k - \mathbf{A}_{\mathrm{ML}}\mathbf{Z}_{\mathrm{ML}})^T \tag{6}$$

Therefore, we arrive at an iterative parameter estimation algorithm by alternating the estimation of $\mathbf{A}$ and $\mathbf{Z}$ via solving Problem (5) and the estimation of $\mathbf{\Psi}$ using (6). The algorithm can be initialized by setting $\mathbf{\Psi} = \frac{1}{KN}\sum_{k=1}^{K}(\mathbf{X}_k - \bar{\mathbf{X}})(\mathbf{X}_k - \bar{\mathbf{X}})^T$. Convergence can be checked by evaluating the log-likelihood function in (3) after each iteration: a very small difference between the values of the log-likelihood function in consecutive iterations indicates the convergence of the algorithm. Moreover, the ERP source number can be determined using the Akaike information criterion (AIC) [9].

For ease of reference, in the following the above algorithm is referred to as *SIgnal-to-noise ratio Maximizer for event-related potentials* (SIM).

III. Experimental Evaluation and Results

The SIM algorithm is evaluated on both simulated data and real EEG recordings. In the simulation where the ground truth is known, the goal is to validate whether the algorithm is able to recover reliably the true model settings generating the simulated data. In the analysis of a P300 EEG data set, the ERP estimates from the algorithm are *qualitatively* assessed according to existing neurophysiological knowledge

[1]The proof of the theorem is omitted due to space limit.

[2]It is assumed that the degenerate case where the correlation matrix of the obtained $\mathbf{Z}_{\mathrm{ML}}$ is isotropic will not occur.

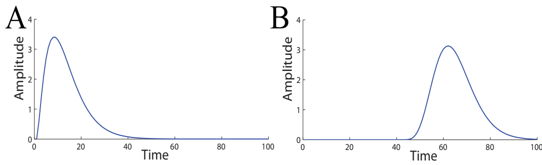

Fig. 1. Simulated ERP sources. The two sources represent the two subcomponents (P100 and P300) of a typical ERP waveform.

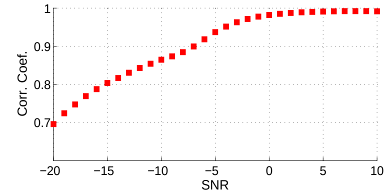

Fig. 2. Run-averaged correlation coefficients between the estimated and true ERP sources for SNRs within $-20 \sim 10$ dB.

about the spatio-temporal properties of ERPs. Moreover, the algorithm is employed for spatial filter design in a target detection task; the classification accuracy serves as a *quantitive* measure for the algorithm's performance in real EEG recordings.

As a comparison to show performance gain, in the target detection task the SIM algorithm is benchmarked against the *regularized SOBI* algorithm [7], as well as an approach using multiple channels without spatial filtering. All computations are done using MATLAB (The MathWorks, Inc.).

A. Simulations

1) Data Description: We investigate SIM's performance at varying SNR levels (-20∼10 dB). The SNR is defined as the power ratio of the overall ERP activities to the overall spontaneous activities in the channel space.

For each SNR, the Monte Carlo simulations consist of 50 runs. In each run, 100 trials of EEG signals are randomly generated. Within each trial, 20 channels of EEG signals are generated as the sum of the linear mixture of 2 ERP sources and uncorrelated noise activities, with each channel comprising 100 data points. The two ERP sources, simulated using the Gamma function, are designed to resemble the two subcomponents (P100 and P300) of a typical ERP waveform (see Fig. 1). The noise activities are simulated as the sum of the linear mixture of 16 spontaneous sources and Gaussian white noise. The time courses of the spontaneous sources are simulated as $1/f$ noise. The variance of the additive noise at each channel is 1/5 of that of the noiseless signal at the same channel. The 20×2 mixing matrix $\mathbf{A}$ for the ERP sources and the 20×16 mixing matrix $\mathbf{B}$ for the spontaneous sources are also randomly generated, with each entry uniformly distributed within $[0, 1]$. The columns of $\mathbf{A}$ are subsequently orthonormalized through Gram-Schmidt process.

2) Results: We apply the SIM algorithm to the simulated data sets. The correlation coefficient between the resulting two estimated ERP sources and the true ones is used as the performance index.

Fig. 2 shows the run-averaged correlation coefficients for all SNR levels. It can be seen that even under poor SNRs (< -10 dB), the recovered ERP sources by SIM show a fairly good match with the true ones, with the correlation coefficients higher than 0.7. As desired asymptotically, when the SNR exceeds 5 dB the correlation coefficient approaches 1, indicating a perfect match. The simulation thus verifies that the SIM algorithm successfully yields accurate estimates for the parameters in model (1).

Furthermore, by tracking the running of the algorithm we find that the increment of the log-likelihood function is negligible after two iterations. Thus it is our belief that the SIM algorithm typically enjoys a high convergence rate.

B. Real EEG Recordings

1) Data Description: Ten subjects (six male and four female, aged 20-28) participated in our *P300-speller* experiments [10]. The speller interface was composed of 36 virtual buttons, each representing a letter or a digit, in the organization of six rows by six columns. During one acquisition period of 15 trials (one block), subjects were instructed to attend to a specific button and mentally count the number of times the row, or the column, containing the designated target character was intensified. Each trial consisted of 12 epochs, each associated with the intensification of a specific row or column specified by a random sequence.

The EEG data were recorded using a Neuroscan SynAmps system. A total of 30 surface electrodes were placed at positions according to the 10/20 international system. Signals were sampled at 200 Hz. A linked-mastoids reference was used. For each subject, single-trial EEG epochs (360 target/1,800 non-target) were derived in association with each stimulus, beginning 200 ms prior to the stimulus onset and lasting for 1,200 ms. All epochs were baseline corrected with respect to the mean voltage over the 200 ms preceding stimulus-onset, and digitally filtered at 1-15 Hz to minimize DC drifts and power line interference.

Determining the presence of the P300 component can be viewed as a *binary classification problem*. For this purpose, each individual data set was split into a training set (180 target/180 nontarget) and a test set (180 target/1,620 nontarget).

Six feature extraction approaches were considered. The 1st approach employed SIM to design spatial filters using all 30-channel EEG from the *target* epochs in the training set, and defined the feature vector as the concatenation of the amplitude of the three estimated ERP sources with the largest power. The number of the ERP sources in the model is determined using AIC. The 2nd∼5th approaches were identical to the first except that regularized SOBI was used for spatial filter design with varying values of the regularization parameter in $\{0, 0.2, 0.6, 1\}$ (0 corresponds to unregularized SOBI). The 6th approach defined the feature vector by simply concatenating the EEG amplitude of six channels, namely Fz, Cz, Pz, Oz, PO7, and PO8. The choice of these six channels for classification was advocated in [11].

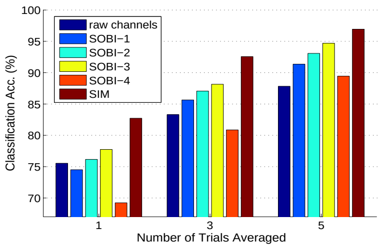

Fig. 3. Classification accuracies of the six feature extraction approaches. SOBI-1∼4 denote regularized SOBI with the value of the regularization parameter being 0, 0.2, 0.6, and 1, respectively.

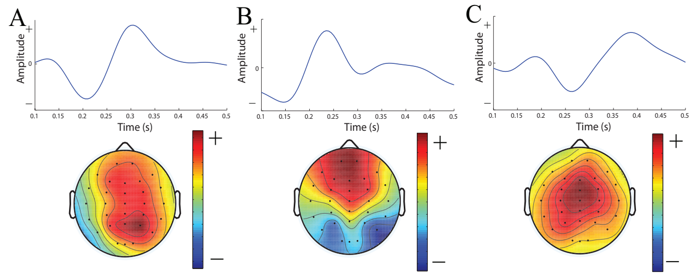

Fig. 4. Spatio-temporal patterns of three estimated ERP sources from one subject. The upper panels show the time courses of the ERP sources (units of the amplitude arbitrary), and the lower panels show the corresponding spatial patterns.

To reduce dimensionality, each feature vector was resampled at 20 Hz.

Fisher discriminant analysis was subsequently employed for feature classification. The classification accuracies on the test sets were used as the performance index for each approach. Due to the low SNR of ERPs, trial average was often performed to improve the classification accuracy. We investigated the effects of number of trials averaged (1, 3, and 5) on the classification performance.

2) Results: Fig. 3 depicts for the six feature extraction approaches the classification accuracies averaged over the ten subjects. It is evident that under the three trial-averaging cases, SIM consistently yields the highest accuracies among all approaches, followed by regularized SOBI with regularization parameter being 0.6 (SOBI-3 in Fig. 3). The improvement is particularly conspicuous when the number of trials averaged is low (5% and 4.42% increase over SOBI-3 for the single-trial and 3-trial averaging cases). The superiority of SIM over the other approaches is highly significant based on the paired-sample Wilcoxon signed rank test (all p-values$<$ 0.005). It is also noteworthy that the single-trial accuracy by SIM is comparable with the 3-trial averaging accuracy by using the six channels without spatial filtering. The excellent results of SIM are not surprising since it has been shown in Section II that the algorithm specifically maximizes the SNR of the ERP sources.

To gain intuition by visualization, Fig. 4 shows the spatio-temporal patterns of the three ERP sources derived using SIM for one subject. According to previous neurophysiology studies, the sources depicted in the left and middle panels could be related to the N2 and P2 peaks, which are the most robust components for the flash VEPs [12][13]; while the sources in the right panel may reflect late visual processing (e.g., P300) [1]. The ERP sources of the other subjects exhibit similar meaningful spatio-temporal patterns (data not shown due to space limit).

IV. Discussion and Conclusion

The proposed spatio-temporal model of ERPs provides a principled framework for further methodological developments. For example, the inter-trial variability in amplitude or latency can be readily taken into consideration in our model by modifying the model structure (e.g., by modeling the variations as random effects in the model [14]). Another realm where our modeling framework may prove to be a valuable starting point is the joint analysis of multi-way (e.g., multi-subject/condition) ERP data. Furthermore, it is worth noting that although our ERP model is introduced in the context of EEG analysis, it should be equally applicable to the analysis of magnetoencephalographic (MEG) data.

In summary, we have proposed a spatio-temporal model with a fast iterative algorithm for estimating ERPs. The promising results suggest that it can be utilized as an effective tool for spatio-temporal analysis of ERPs as well as spatial filter design for decoding brain states.

References

[1] S. L. Bressler, "Event-related potentials," in: *The Handbook of Brain Theory and Neural Networks*, MIT Press, Cambridge MA, pp. 412-415, 2002.

[2] B. Blankertz et al., "Single-trial analysis and classification of ERP components – a tutorial," *Neuroimage*, doi: 10.1016/j.neuroimage.2010.06.048, 2011.

[3] A. Hyvärinen et al., "Independent component analysis of short-time Fourier transforms for spontaneous EEG/MEG analysis," *Neuroimage*, vol. 49, pp. 257-271, 2010.

[4] R. M. Chapman et al., "EP component identification and measurement by principal component analysis," *Brain Cogn.*, vol. 27, pp. 288-310, 1995.

[5] S. Makeig et al., "Independent component analysis of electroencephalographic data," *Adv. Neural Inf. Proc. Syst.*, vol. 8, pp. 145-151, 1996.

[6] M. Zibulevsky et al., "Extraction of a single source from multichannel data using sparse decomposition," *Neurocomputing*, vol. 49, pp. 163-173, 2002.

[7] S. Lemm et al., "Enhancing the signal-to-noise ratio of ICA-based extracted ERPs," *IEEE Transactions on Biomedical Engineering*, vol. 53, pp. 601-607, 2006.

[8] R. Li et al., "A spatiotemporal filtering methodology for single-trial ERP component estimation," *IEEE Transactions on Biomedical Engineering*, vol. 56, pp. 83-92, 2009.

[9] Y. Pawitan, *In All Likelihood: Statistical Modelling and Inference Using Likelihood*, Oxford University Press, 2001.

[10] B. Hong et al., "N200-speller using motion-onset visual response," *Clinical Neurophysiology*, vol. 120, pp. 1658-1666, 2009.

[11] D. J. Krusienski et al., "Towards enhanced P300 speller performance," *J Neurosci Methods*, vol. 167, pp. 15-21, 2007.

[12] M. M. Müller et al., "Concurrent recording of steady-state and transient event-related potentials as indices of visual spatial selective attention," *Clinical Neurophysiology*, vol. 111, pp. 1544-1552, 2000.

[13] J. Odom et al., "Visual evoked potentials standard," *Documenta Ophthalmologica*, vol. 108, pp. 115-123, 2004.

[14] W. Wu et al., "A hierarchical Bayesian approach for learning sparse spatio-temporal decompositions of multichannel EEG," *Neuroimage*, vol. 56, pp. 1929-1945, 2011.